



Advanced Quantitative Methods: Ordinary Least Squares

Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

Johan A. Elkind

University College Dublin

2 February 2018



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

- 1 Linear model
- 2 Ordinary Least Squares

Assumptions

Algebra

Properties

- 3 Testing in regression

Probability distributions

t - and F -tests



Outline

- 1 Linear model
- 2 Ordinary Least Squares
 - Assumptions
 - Algebra
 - Properties
- 3 Testing in regression
 - Probability distributions
 - t - and F -tests

Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests



Terminology

- \mathbf{y} is the **dependent** variable
 - referred to also (by Greene) as a **regressand**
- \mathbf{X} are the **independent** variables
 - also known as **explanatory** variables
 - also known as **regressors** or **predictors** (or **factors**, **carriers**)
 - \mathbf{X} is sometimes called the **design matrix** (or **factor space**)
- \mathbf{y} is **regressed on** \mathbf{X}
- The **error** term ε is sometimes called a **disturbance**:
$$\varepsilon = \mathbf{y} - \mathbf{X}\beta.$$
- The difference between the observed and predicted dependent variable is called the **residual**:
$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta}.$$



$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon_i$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2)$$

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2)$$



Components

Two components of the model:

$$\begin{array}{l|l} \mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2) & \text{Stochastic} \\ \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} & \text{Systematic} \end{array}$$

Generalised version (not necessarily linear):

$$\begin{array}{l|l} \mathbf{y} \sim f(\boldsymbol{\mu}, \boldsymbol{\alpha}) & \text{Stochastic} \\ \boldsymbol{\mu} = g(\mathbf{X}, \boldsymbol{\beta}) & \text{Systematic} \end{array}$$



Components

$$\begin{array}{l|l} \mathbf{y} \sim f(\boldsymbol{\mu}, \boldsymbol{\alpha}) & \text{Stochastic} \\ \boldsymbol{\mu} = g(\mathbf{X}, \boldsymbol{\beta}) & \text{Systematic} \end{array}$$

Stochastic component: varies over repeated (hypothetical) observations on the same unit.

Systematic component: varies across units, but constant given **X**.

(King, 1998, 8)



Uncertainty

$$\begin{array}{l|l} \mathbf{y} \sim f(\boldsymbol{\mu}, \boldsymbol{\alpha}) & \text{Stochastic} \\ \boldsymbol{\mu} = g(\mathbf{X}, \boldsymbol{\beta}) & \text{Systematic} \end{array}$$

Two types of uncertainty:

Estimation uncertainty: lack of knowledge about $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$; can be reduced by increasing n .

Fundamental uncertainty: represented by stochastic component and exists independent of researcher.



Outline

- 1 Linear model
- 2 Ordinary Least Squares
 - Assumptions
 - Algebra
 - Properties
- 3 Testing in regression
 - Probability distributions
 - t - and F -tests

Ordinary Least Squares (OLS)



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

For the **linear** model, the most popular method of estimation is **ordinary least squares** (OLS).

$\hat{\beta}^{OLS}$ are those estimates of β that minimize the sum of squared residuals: $\mathbf{e}'\mathbf{e}$.

OLS is the **best linear unbiased estimator** (BLUE).



Outline

1 Linear model

2 Ordinary Least Squares

Assumptions

Algebra

Properties

3 Testing in regression

Probability distributions

t - and F -tests

Assumptions: specification



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

- Linear in parameters (i.e. $f(\mathbf{X}\beta) = \mathbf{X}\beta$ and $E(\mathbf{y}) = \mathbf{X}\beta$)
- No extraneous variables in \mathbf{X}
- No omitted independent variables
- Parameters to be estimated are constant
- Number of parameters is less than the number of cases,
 $k < n$

Assumptions: errors



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

- Errors have an expected value of zero, $E(\varepsilon|\mathbf{X}) = 0$
- Errors are normally distributed, $\varepsilon \sim N(0, \sigma^2)$
- Errors have a constant variance, $\text{var}(\varepsilon|\mathbf{X}) = \sigma^2 < \infty$
- Errors are not autocorrelated, $\text{cov}(\varepsilon_i, \varepsilon_j|\mathbf{X}) = 0 \quad \forall \quad i \neq j$
- Errors and \mathbf{X} are uncorrelated, $\text{cov}(\mathbf{X}, \varepsilon) = 0$

Assumptions: regressors



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

- \mathbf{X} varies
- \mathbf{X} is of full column rank (note: requires $k < n$)
- No measurement error in \mathbf{X}
- No endogenous variables in \mathbf{X}



Outline

- 1 Linear model
- 2 Ordinary Least Squares

Assumptions

Algebra

Properties

- 3 Testing in regression

Probability distributions

t - and F -tests

Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$



$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}_{n \times k} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{k \times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} \beta_1 + \beta_2 x_{12} + \beta_3 x_{13} + \cdots + \beta_k x_{1k} \\ \beta_1 + \beta_2 x_{22} + \beta_3 x_{23} + \cdots + \beta_k x_{2k} \\ \vdots \\ \beta_1 + \beta_2 x_{n2} + \beta_3 x_{n3} + \cdots + \beta_k x_{nk} \end{bmatrix}_{n \times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}$$



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}$$

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{OLS} &= \arg \min_{\hat{\boldsymbol{\beta}}} \mathbf{e}'\mathbf{e} \\ &= \arg \min_{\hat{\boldsymbol{\beta}}} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\end{aligned}$$



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

$$\begin{aligned} \mathbf{e}'\mathbf{e} &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= (\mathbf{y}' - (\mathbf{X}\hat{\beta})')(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{y}'\mathbf{y} - (\mathbf{X}\hat{\beta})'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} + (\mathbf{X}\hat{\beta})'\mathbf{X}\hat{\beta} \\ &= \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \end{aligned}$$

Deriving $\hat{\beta}^{OLS}$



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

$$\hat{\beta}^{OLS} = \arg \min_{\hat{\beta}} \mathbf{e}'\mathbf{e} \implies$$

$$\frac{\partial(\mathbf{e}'\mathbf{e})}{\partial \hat{\beta}^{OLS}} = 0$$

$$\frac{\partial(\mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta})}{\partial \hat{\beta}^{OLS}} = 0$$



Deriving $\hat{\beta}^{OLS}$

$$\frac{\partial(\mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta})}{\partial\hat{\beta}^{OLS}} = 0$$

$$2\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} - 2\mathbf{X}'\mathbf{y} = 0$$

$$2\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = 2\mathbf{X}'\mathbf{y}$$

$$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$$

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$\hat{\beta}^{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Note that $\hat{\beta}'\mathbf{X}'\mathbf{y}$ is a scalar, so we can take the transpose at will. We make use of this in the derivative, where $\partial((\mathbf{y}'\mathbf{X}\hat{\beta})')/\partial\hat{\beta} = \mathbf{X}'\mathbf{y}$ (Harville, 2008, 300).

$$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$$



$$\begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}'_{n \times k} \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}_{n \times k} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}_k$$

$$= \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}'_{n \times k} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_n$$

($\hat{\beta}$ here refers to OLS estimates.)

$$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$$



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{12} & x_{22} & \cdots & x_{n2} \\ x_{13} & x_{23} & \cdots & x_{n3} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix}_{k \times n} \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}_{n \times k} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}_{k \times 1} \\
 = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{12} & x_{22} & \cdots & x_{n2} \\ x_{13} & x_{23} & \cdots & x_{n3} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix}_{k \times n} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$

$$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$$



$$\begin{bmatrix} n & \sum x_{i2} & \cdots & \sum x_{ik} \\ \sum x_{i2} & \sum (x_{i2})^2 & \cdots & \sum x_{i2}x_{ik} \\ \sum x_{i3} & \sum x_{i3}x_{i2} & \cdots & \sum x_{i3}x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_{ik} & \sum x_{ik}x_{i2} & \cdots & \sum (x_{ik})^2 \end{bmatrix}_{k \times k} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}_{k \times 1} = \begin{bmatrix} \sum y_i \\ \sum x_{i2}y_i \\ \sum x_{i3}y_i \\ \vdots \\ \sum x_{ik}y_i \end{bmatrix}_{k \times 1}$$

(\sum refers to \sum_i^n)

$$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$$

So this can be seen as a set of linear equations to solve:

$$\begin{aligned} \hat{\beta}_1 n + \hat{\beta}_2 \sum x_{i2} + \cdots + \hat{\beta}_k \sum x_{ik} &= \sum y_i \\ \hat{\beta}_1 \sum x_{i2} + \hat{\beta}_2 \sum (x_{i2})^2 + \cdots + \hat{\beta}_k \sum x_{i2}x_{ik} &= \sum x_{i2}y_i \\ \hat{\beta}_1 \sum x_{i3} + \hat{\beta}_2 \sum x_{i3}x_{i2} + \cdots + \hat{\beta}_k \sum x_{i3}x_{ik} &= \sum x_{i3}y_i \\ &\vdots \\ \hat{\beta}_1 \sum x_{ik} + \hat{\beta}_2 \sum x_{ik}x_{i2} + \cdots + \hat{\beta}_k \sum (x_{ik})^2 &= \sum x_{ik}y_i \end{aligned}$$



Deriving $\text{var}(\hat{\beta}^{OLS})$



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

$$\begin{aligned}\text{var}(\hat{\beta}^{OLS}) &= E[(\hat{\beta}^{OLS} - E(\hat{\beta}^{OLS}))(\hat{\beta}^{OLS} - E(\hat{\beta}^{OLS}))'] \\ &= E[(\hat{\beta}^{OLS} - \beta)(\hat{\beta}^{OLS} - \beta)']\end{aligned}$$

Deriving $var(\hat{\beta}^{OLS})$



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

$$\begin{aligned}\hat{\beta}^{OLS} - \beta &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \beta \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon) - \beta \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon - \beta \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon - \beta \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\end{aligned}$$

Deriving $var(\hat{\beta}^{OLS})$



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

$$\begin{aligned}var(\hat{\beta}^{OLS}) &= E[(\hat{\beta}^{OLS} - \beta)(\hat{\beta}^{OLS} - \beta)'] \\&= E[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon)((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon)'] \\&= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\varepsilon\varepsilon']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$



Estimating σ^2

$$\begin{aligned} \mathbf{e} &= \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{OLS} \\ &= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \\ &= \mathbf{M}\mathbf{y} \\ &= \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\varepsilon} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\varepsilon} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\varepsilon} \\ &= \mathbf{M}\boldsymbol{\varepsilon} \end{aligned}$$

Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

Estimating σ^2



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

 t - and F -tests

$$\begin{aligned} \mathbf{e}'\mathbf{e} &= (\mathbf{M}\boldsymbol{\varepsilon})'(\mathbf{M}\boldsymbol{\varepsilon}) \\ &= \boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{M}\boldsymbol{\varepsilon} \\ &= \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon} \end{aligned}$$

$$E[\mathbf{e}'\mathbf{e}|\mathbf{X}] = E[\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}|\mathbf{X}]$$

$\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$ is a scalar, so if you see it as a 1×1 matrix, it is equal to its trace, therefore

$$\begin{aligned} E[\mathbf{e}'\mathbf{e}|\mathbf{X}] &= E[\text{tr}(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon})|\mathbf{X}] \\ &= E[\text{tr}(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')|\mathbf{X}] \\ &= \text{tr}(\mathbf{M}E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}]) \\ &= \text{tr}(\mathbf{M}\sigma^2\mathbf{I}) = \sigma^2 \text{tr}(\mathbf{M}) \end{aligned}$$



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

$$\begin{aligned}tr(\mathbf{M}) &= tr(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\ &= tr(\mathbf{I}) - tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\ &= n - tr(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) \\ &= n - k\end{aligned}$$

$$E(\mathbf{e}'\mathbf{e}|\mathbf{X}) = (n - k)\sigma^2$$

$$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n - k}$$



Standard errors

$$\text{var}(\hat{\beta}^{OLS}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

The standard errors of $\hat{\beta}^{OLS}$ are then the square root of the diagonal of this matrix.

In the simple case where $\mathbf{y} = \beta_0 + \beta_1\mathbf{x}$, this gives

$$\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i^n (x_i - \bar{x})^2}$$

Note how an increase in variance in \mathbf{x} leads a decrease in the standard error of $\hat{\beta}_1$.



Outline

1 Linear model

2 Ordinary Least Squares

Assumptions

Algebra

Properties

3 Testing in regression

Probability distributions

t - and F -tests

Unbiasedness of $\hat{\beta}^{OLS}$

From deriving $var(\hat{\beta}^{OLS})$ we have:

$$\begin{aligned}\hat{\beta}^{OLS} - \beta &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \beta \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon) - \beta \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon - \beta \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon - \beta \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\end{aligned}$$

Since we assume $E(\varepsilon) = 0$:

$$E(\hat{\beta}^{OLS} - \beta) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\varepsilon) = 0$$

Therefore, $\hat{\beta}^{OLS}$ is an **unbiased** estimator of β .



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

Consistency of $\hat{\beta}^{OLS}$

Consistency requires that both the bias and the variance of $\hat{\beta}^{OLS}$ approach zero as $n \rightarrow \infty$.

Since $\hat{\beta}^{OLS}$ is unbiased in a finite sample, it is also unbiased as n increases.

$$\text{var}(\hat{\beta}^{OLS}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \frac{\sigma^2}{n} \left(\frac{1}{n} \mathbf{X}'\mathbf{X} \right)^{-1}$$

$\mathbf{X}'\mathbf{X}$ would typically increase with n , but if we can assume that $\frac{1}{n}\mathbf{X}'\mathbf{X}$ converges to a finite, invertible matrix \mathbf{Q} ,

$$\lim_{n \rightarrow \infty} \text{var}(\hat{\beta}^{OLS}) = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} \mathbf{Q}^{-1} = \mathbf{0}.$$





Consistency of $\hat{\beta}^{OLS}$

An alternative approach to deriving the consistency of $\hat{\beta}^{OLS}$:

$$\begin{aligned}\hat{\beta}^{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon) \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon \\ &= \beta + \left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1} \left(\frac{1}{n}\mathbf{X}'\varepsilon\right)\end{aligned}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n}\mathbf{X}'\mathbf{X} = \mathbf{Q} \implies \lim_{n \rightarrow \infty} \left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1} = \mathbf{Q}^{-1}$$

$$E \left[\frac{1}{n}\mathbf{X}'\varepsilon \right] = E \left[\frac{1}{n} \sum_i^n \mathbf{x}_i \varepsilon_i \right] = 0$$



$$\begin{aligned} \text{var} \left(\frac{1}{n} \mathbf{X}' \boldsymbol{\varepsilon} \right) &= E \left[\frac{1}{n} \mathbf{X}' \boldsymbol{\varepsilon} \left(\frac{1}{n} \mathbf{X}' \boldsymbol{\varepsilon} \right)' \right] \\ &= \frac{1}{n} \mathbf{X}' E[\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}'] \mathbf{X} \frac{1}{n} \\ &= \left(\frac{\sigma^2}{n} \right) \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right) \end{aligned}$$

$$\lim_{n \rightarrow \infty} \text{var} \left(\frac{1}{n} \mathbf{X}' \boldsymbol{\varepsilon} \right) = \mathbf{0Q} = \mathbf{0}$$

Consistency of $\hat{\beta}^{OLS}$



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

$$\hat{\beta}^{OLS} = \beta + \left(\frac{1}{n} \mathbf{X}'\mathbf{X} \right)^{-1} \left(\frac{1}{n} \mathbf{X}'\varepsilon \right)$$

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}'\mathbf{X} \right)^{-1} = \mathbf{Q}^{-1}$$

$$E \left[\frac{1}{n} \mathbf{X}'\varepsilon \right] = 0$$

$$\lim_{n \rightarrow \infty} \text{var} \left(\frac{1}{n} \mathbf{X}'\varepsilon \right) = 0$$

imply that the sampling distribution of $\hat{\beta}^{OLS}$ “collapses” to β as n becomes very large, i.e.:

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}^{OLS} = \beta$$



Efficiency of $\hat{\beta}^{OLS}$

The **Gauss-Markov theorem** states that there is no **linear unbiased estimator** of β that has a smaller sampling variance than $\hat{\beta}^{OLS}$, i.e. $\hat{\beta}^{OLS}$ is BLUE.

An estimator is **linear** iff it can be expressed as a linear function of the data on the dependent variable:
 $\hat{\beta}_j^{linear} = \sum_i^n f(x_{ij})y_i$, which is the case for OLS:

$$\begin{aligned}\hat{\beta}^{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{C}\mathbf{y}\end{aligned}$$



Gauss-Markov Theorem

$$\begin{aligned}\hat{\beta}^{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{C}\mathbf{y}\end{aligned}$$

Imagine we have another linear estimator: $\hat{\beta}^* = \mathbf{W}\mathbf{y}$, where $\mathbf{W} = \mathbf{C} + \mathbf{D}$.

$$\begin{aligned}\hat{\beta}^* &= (\mathbf{C} + \mathbf{D})\mathbf{y} \\ &= \mathbf{C}\mathbf{y} + \mathbf{D}\mathbf{y} \\ &= \hat{\beta}^{OLS} + \mathbf{D}(\mathbf{X}\beta + \varepsilon) \\ &= \hat{\beta}^{OLS} + \mathbf{D}\mathbf{X}\beta + \mathbf{D}\varepsilon\end{aligned}$$

Gauss-Markov Theorem



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

$$\begin{aligned} E(\hat{\beta}^*) &= E \left[\hat{\beta}^{OLS} + \mathbf{DX}\beta + \mathbf{D}\varepsilon \right] \\ &= E \left[\hat{\beta}^{OLS} \right] + \mathbf{DX}\beta + E \left[\mathbf{D}\varepsilon \right] \\ &= \beta + \mathbf{DX}\beta + 0 \\ \Rightarrow \mathbf{DX}\beta &= 0 \\ \Rightarrow \mathbf{DX} &= 0, \end{aligned}$$

because both $\hat{\beta}^{OLS}$ and $\hat{\beta}^*$ are unbiased, so
 $E(\hat{\beta}^{OLS}) = E(\hat{\beta}^*) = \beta$.

Gauss-Markov Theorem



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

$$\begin{aligned}\hat{\beta}^* &= \hat{\beta}^{OLS} + \mathbf{D}\varepsilon \\ \hat{\beta}^* - \beta &= \hat{\beta}^{OLS} + \mathbf{D}\varepsilon - \beta \\ &= (\mathbf{C} + \mathbf{D})\varepsilon \\ \text{var}(\hat{\beta}^*) &= E(\hat{\beta}^* - \beta)E(\hat{\beta}^* - \beta)' \\ &= E((\mathbf{C} + \mathbf{D})\varepsilon)E((\mathbf{C} + \mathbf{D})\varepsilon)' \\ &= (\mathbf{C} + \mathbf{D})E(\varepsilon\varepsilon')(\mathbf{C} + \mathbf{D})' \\ &= \sigma^2(\mathbf{C} + \mathbf{D})(\mathbf{C} + \mathbf{D})' \\ &= \sigma^2(\mathbf{C}\mathbf{C}' + \mathbf{D}\mathbf{C}' + \mathbf{C}\mathbf{D}' + \mathbf{D}\mathbf{D}') \\ &= \sigma^2((\mathbf{X}'\mathbf{X})^{-1} + \mathbf{D}\mathbf{D}') \\ &\geq \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Sums of squares



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

SST Total sum of squares $\sum (y_i - \bar{y})^2$

SSE Explained sum of squares $\sum (\hat{y}_i - \bar{y})^2$

SSR Residual sum of squares

$$\sum e_i^2 = \sum (\hat{y}_i - y_i)^2 = \mathbf{e}'\mathbf{e}$$

The key to remember is that **SST = SSE + SSR**

Sometimes instead of “explained” and “residual”, “regression” and “error” are used, respectively, so that the abbreviations are swapped (!).



Defined in terms of sums of squares:

$$\begin{aligned}R^2 &= \frac{SSE}{SST} \\ &= 1 - \frac{SSR}{SST} \\ &= 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}\end{aligned}$$

Interpretation: the proportion of the variation in \mathbf{y} that is explained linearly by the independent variables.

A much over-used statistic: it may not be what we are interested in at all!



When a model has no intercept, it is possible for R^2 to lie outside the interval $(0, 1)$.

R^2 rises with the addition of more explanatory variables. For this reason we often report the **adjusted** R^2 :

$$1 - (1 - R^2) \frac{n - 1}{n - k}.$$

The R^2 values from different \mathbf{y} samples *cannot be compared*.



Outline

- 1 Linear model
- 2 Ordinary Least Squares
 - Assumptions
 - Algebra
 - Properties
- 3 Testing in regression
 - Probability distributions
 - t - and F -tests



Outline

1 Linear model

2 Ordinary Least Squares

Assumptions

Algebra

Properties

3 Testing in regression

Probability distributions

t - and F -tests



Bernoulli trial

“An experiment in which s trials are made of an event, with probability p of success in any given trial.”

(Weisstein, Eric W. “Bernoulli Trial.” <http://mathworld.wolfram.com/BernoulliTrial.html>)



Binomial distribution

“The (...) probability distribution (...) of obtaining exactly n successes out of N Bernoulli trials.”

$$P(n|N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

(Weisstein, Eric W. “Binomial Distribution.” <http://mathworld.wolfram.com/BinomialDistribution.html>)



Binomial distribution

$$P(n|N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$
$$\lim_{N \rightarrow \infty} p(n) = \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(n-Np)^2}{2Npq}}$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(n-\bar{n})^2}{2\sigma^2}}$$

i.e. the **limiting distribution** of the binomial distribution is the **normal distribution**, with $\sigma^2 \equiv Npq$.

(Weisstein, Eric W. "Binomial distribution." <http://mathworld.wolfram.com/binomialdistribution.html>)



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

Normal distribution

Also called **Gaussian distribution** , but Gauss did not invent it.

(Davidson and MacKinnon, 1999, 130–135)



The sum of squares of r independent standard normal distributions, is distributed chi-squared with r degrees of freedom, i.e. if $x \sim N(0, 1)$, then:

$$\sum_i^r x_i^2 \sim \chi^2(r)$$

(Weisstein, Eric W. "Chi-squared distribution." <http://mathworld.wolfram.com/chi-squareddistribution.html>)

χ^2 -distribution



Outline

Linear model

OLS

Assumptions

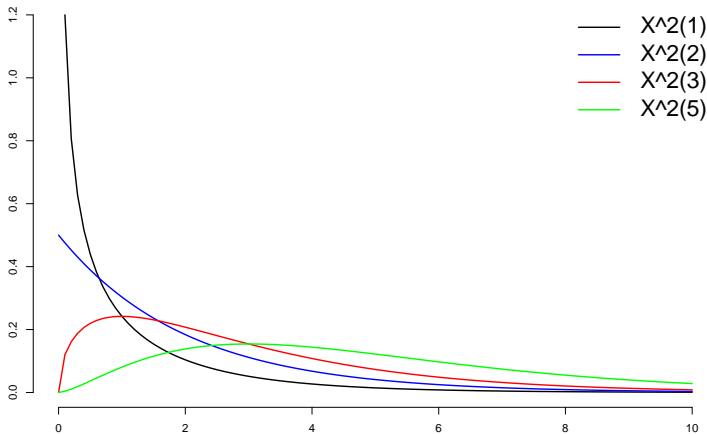
Algebra

Properties

Testing

Distributions

t - and F -tests





F -distribution

If

$$x \sim \chi^2(m)$$

$$y \sim \chi^2(n)$$

with x , y independent , then

$$\frac{x/m}{y/n} \sim F(m, n)$$

has an F -distribution with m and n degrees of freedom.

F-distribution



Outline

Linear model

OLS

Assumptions

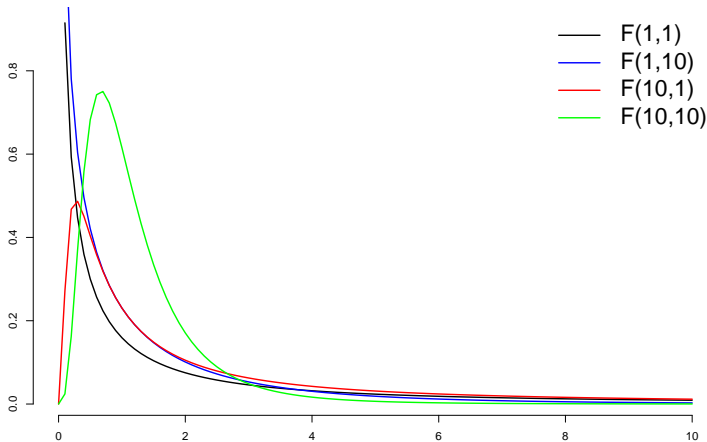
Algebra

Properties

Testing

Distributions

t - and F -tests





t -distribution

If

$$x \sim N(0, 1)$$

$$y \sim \chi^2(r)$$

with x , y independent , then

$$\frac{x}{\sqrt{y/r}} \sim t(r)$$

has a t -distribution with r degrees of freedom.



t -distribution

Imagine we have a sample of size n and we calculate the sample mean of x :

$$\bar{x} = \frac{1}{n} \sum_i^n x_i$$

with variance estimator

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_i^n (x_i - \bar{x})^2$$

then $\hat{\sigma}^2 \sim \chi^2(n-1)$ (because $\hat{\sigma}^2$ is a sum of squares and the use of deviations from the mean removes one degree of freedom).



t -distribution

\bar{x} is a sum of normally distributed values, so is itself normally distributed; $\hat{\sigma}^2$ has a χ^2 distribution, so

$$t(n) = \frac{\bar{x} - \mu}{\sqrt{\hat{\sigma}^2/n}}$$

has a t -distribution with $n - 1$ degrees of freedom.

As n increases, the t -distribution approaches a normal distribution.

The t -distribution is the approximation for the normal distribution when n is small and σ^2 unknown.

(The $t(1)$ distribution is also called the **Cauchy distribution**.)

t -distribution



Outline

Linear model

OLS

Assumptions

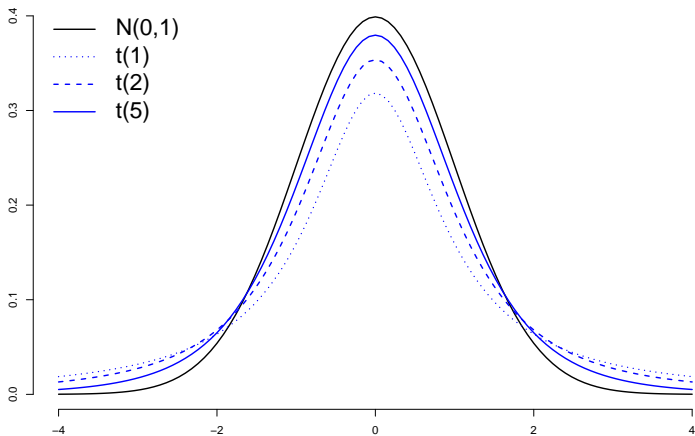
Algebra

Properties

Testing

Distributions

t - and F -tests





Outline

- 1 Linear model
- 2 Ordinary Least Squares
 - Assumptions
 - Algebra
 - Properties
- 3 Testing in regression
 - Probability distributions
 - t - and F -tests

Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests



t -test

$$h_0 : \beta = 0$$

$$h_1 : \beta \neq 0$$

We can calculate the t -value by subtracting the value under the null and dividing by the standard error:

$$t = \frac{\hat{\beta}}{\sigma_{\hat{\beta}}}$$

Because $\hat{\beta}$ has a normal distribution and $\sigma_{\hat{\beta}}^2$ a χ^2 -distribution, t has the t -distribution with $n - k$ degrees of freedom.



t-test

Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t- and *F*-tests

$$h_0 : \beta = 0$$

$$h_1 : \beta \neq 0$$

$$t = \frac{\hat{\beta}}{\sigma_{\hat{\beta}}}$$

$$h_0 : \beta = a$$

$$h_1 : \beta \neq a$$

$$t = \frac{\hat{\beta} - a}{\sigma_{\hat{\beta}}}$$

$$h_0 : \beta_3 = \beta_2$$

$$h_1 : \beta_3 \neq \beta_2$$

$$t = \frac{\hat{\beta}_3 - \hat{\beta}_2}{\sigma_{\hat{\beta}_3}}$$

Sums of squares



SST Total sum of squares $\sum (y_i - \bar{y})^2$

SSE Explained sum of squares $\sum (\hat{y}_i - \bar{y})^2$

SSR Residual sum of squares

$$\sum e_i^2 = \sum (\hat{y}_i - y_i)^2 = \mathbf{e}'\mathbf{e}$$

The key to remember is that **SST = SSE + SSR**

Sometimes instead of “explained” and “residual”, “regression” and “error” are used, respectively, so that the abbreviations are swapped (!).



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t- and F-tests

Variation	SS	df	MS
Explained	$\sum(\hat{y}_i - \bar{y})^2$	$k - 1$	$SSE/(k - 1)$
Residual	$\sum(\hat{y}_i - y_i)^2$	$n - k$	$SSR/(n - k)$
Total	$\sum(y_i - \bar{y})^2$	$n - 1$	

$$\frac{SSE/(k - 1)}{SSR/(n - k)} \sim F_{k-1, n-k}$$



Restrictions

$$H_0 : \mathbf{y} = \mathbf{X}_{n \times k}^{(1)} \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$$

$$H_1 : \mathbf{y} = \mathbf{X}_{n \times k}^{(1)} \boldsymbol{\beta}_1 + \mathbf{X}_{n \times r}^{(2)} \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

USSR Unrestricted sum of squared residuals

RSSR Restricted sum of squared residuals

$$F_{\beta_2} = \frac{(RSSR - USSR)/r}{USSR/(n - k - r)} \sim F(r, n - k - r)$$

Chow test



You might have a situation where you want to know whether the coefficients differ for two groups in the sample.

You could test this with the following regression:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \beta + \begin{bmatrix} \mathbf{0} \\ \mathbf{X}_2 \end{bmatrix} \gamma + \varepsilon,$$

whereby the difference between the coefficients for the two groups are captured by γ .



It turns out you can just run two separate regressions (“unrestricted”), for the two groups, as well as (“restricted”) regression \mathbf{y} on \mathbf{X} , so that we get:

$$F_{\gamma} = \frac{(RSSR - SSR_1 - SSR_2)/k}{(SSR_1 + SSR_2)/(n - 2k)} \sim F(k, n - 2k)$$

(Davidson and MacKinnon, 1999, 146–147)



Type I and II errors

Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

Type I error: rejecting a null hypothesis that is true (e.g. $\alpha = .05$)

Type II error: not rejecting a null hypothesis that is false

(Davidson and MacKinnon, 1999, 126)



Power

Power: probability of rejecting a hypothesis that is false

$$1 - P(\text{Type II error})$$

The power of a test increases when:

- the true value is further from the null hypothesis value;
- the variance is lower;
- the sample size is larger.

(Davidson and MacKinnon, 1999, 126)



Outline

Linear model

OLS

Assumptions

Algebra

Properties

Testing

Distributions

t - and F -tests

The p -value is the probability of a Type I error when rejecting the null hypothesis.

You can say a test is “statistically significant” if $p < \alpha$, but the p -value contains more information by itself.

(Davidson and MacKinnon, 1999, 128)

$$\alpha = .05$$



Note that his value is absolutely arbitrary and just habit since the publication of Fisher (1922).

The p -value is, one could argue, just a complicated way of measuring the sample size.

“Another interesting example (...) is the propensity for published studies to contain a disproportionately large number of Type I errors; studies with statistically significant results tend to get published, whereas those with insignificant results do not.” (Kennedy, 2008, 61)

$$\alpha = .05$$



Outline

Linear model

OLS

Assumptions

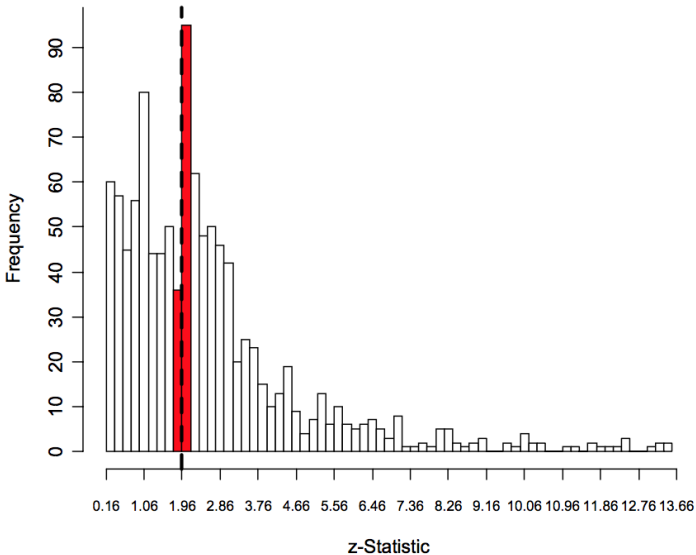
Algebra

Properties

Testing

Distributions

t - and F -tests



ordinary least
squares



Appendix

Probabilities

Probability
distributions
in R

OLS in R
manually

Likelihood
tests

References

Appendix

$$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$$

When there is only one independent variable, this reduces to:

$$\hat{\beta}_1 n + \hat{\beta}_2 \sum x_i = \sum y_i$$

$$\hat{\beta}_1 \sum x_i + \hat{\beta}_2 \sum x_i^2 = \sum x_i y_i$$

$$\hat{\beta}_1 = \frac{\sum y_i - \hat{\beta}_2 \sum x_i}{n} = \frac{n\bar{y} - \hat{\beta}_2 n\bar{x}}{n} = \bar{y} - \hat{\beta}_2 \bar{x}$$

$$(\bar{y} - \hat{\beta}_2 \bar{x}) \sum x_i + \hat{\beta}_2 \sum x_i^2 = \sum x_i y_i$$

$$n\bar{y}\bar{x} - \hat{\beta}_2 n\bar{x}^2 + \hat{\beta}_2 \sum x_i^2 = \sum x_i y_i$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i - n\bar{y}\bar{x}}{\sum x_i^2 - n\bar{x}^2}$$



Simple regression



Appendix

Probabilities

Probability
distributions
in R

OLS in R
manually

Likelihood
tests

References

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

If $\beta_1 = 0$ and the only regressor is the intercept, then $\hat{\beta}_0 = \bar{y}$.

If $\beta_0 = 0$, so that there is no intercept and one explanatory variable \mathbf{x} , then $\hat{\beta}_1 = \frac{\sum_i^n x_i y_i}{\sum_i^n x_i^2}$.

If there is an intercept and one explanatory variable, then

$$\hat{\beta}_1 = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}$$



Appendix

Probabilities

Probability
distributions
in R

OLS in R
manually

Likelihood
tests

References

Demeaning

If the observations are expressed as deviations from their means, i.e.:

$$y_i^* = y_i - \bar{y}$$

$$x_i^* = x_i - \bar{x},$$

then

$$\hat{\beta}_1 = \frac{\sum_i^n x_i^* y_i^*}{\sum_i^n x_i^{*2}}$$

The intercept can be estimated as $\bar{y} - \hat{\beta}_1 \bar{x}$.



Standardized variables

The coefficient on a variable is interpreted as the effect on \mathbf{y} given a one unit increase in \mathbf{x} and thus the interpretation is dependent on the scale of measurement.

An option can be to standardize the variables:

$$\mathbf{y}^* = \frac{\mathbf{y} - \bar{y}}{\sqrt{\sigma_y^2}}$$
$$\mathbf{x}^* = \frac{\mathbf{x} - \bar{x}}{\sqrt{\sigma_x^2}}$$
$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta}^* + \varepsilon^*$$

and interpret the coefficients as the effect on \mathbf{y} expressed in standard deviations as \mathbf{x} increase by one standard deviation.

Standardized variables



Appendix

Probabilities

Probability
distributions
in R

OLS in R
manually

Likelihood
tests

References

```
library(arm)
summary(standardize(lm(y ~ x)),
        binary.inputs="leave.alone")
```

See also Andrew Gelman (2007), "Scaling regression inputs by dividing by two standard deviations".



Outline

- 4 Probabilities
- 5 Probability distributions in R
- 6 OLS in R manually
- 7 Likelihood tests

Appendix

Probabilities

Probability
distributions
in R

OLS in R
manually

Likelihood
tests

References



i.i.d.

We make three assumptions about our data to proceed:

- The observations are **independent**
- The observations are **identically distributed**
- The population has a finite mean and a finite variance

A variable for which the first two assumptions hold is called **iid**.



Independent observations

Intuitively: the value for one case does not affect the value for another case on the same variable.

More formally: $P(x_1 \cap x_2) = P(x_1)P(x_2)$.

Examples of dependent observations:

- grades of students in different classes;
- stock values over time;
- economic growth in neighbouring countries.



Identically distributed

All the observations are drawn from the same **random variable** with the same **probability distribution**.

An example where this is not the case would generally be panel data. E.g. larger firms will have larger variations in profits, thus their variance differs, thus these are not observations from the same probability distribution.



Random sample

A proper **random sample** is i.i.d.

The law of large numbers and the Central Limit Theorem help us to predict the behaviour of our sample data.



Law of large numbers

Appendix

Probabilities

Probability
distributions
in R

OLS in R
manually

Likelihood
tests

References

The law of large numbers (LLN) states that, if these three assumptions are satisfied, the sample mean will approach the population mean with probability one if the sample is infinitely large.

Central Limit Theorem



Appendix

Probabilities

Probability
distributions
in R

OLS in R
manually

Likelihood
tests

References

If these three assumptions are satisfied,

Central Limit Theorem



Appendix

Probabilities

Probability
distributions
in R

OLS in R
manually

Likelihood
tests

References

If these three assumptions are satisfied,

- The sample mean is **normally distributed**, *regardless of the distribution of the original variable.*



Central Limit Theorem

If these three assumptions are satisfied,

- The sample mean is **normally distributed**, *regardless of the distribution of the original variable*.
- The sample mean has the **same expected value** as the population mean (LLN).

Central Limit Theorem



If these three assumptions are satisfied,

- The sample mean is **normally distributed**, *regardless of the distribution of the original variable*.
- The sample mean has the **same expected value** as the population mean (LLN).
- The standard deviation (**standard error**) of the sample mean is: $S.E.(\bar{x}) = \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$.



Sample and population size

Note that the standard error depends only on the sample size,
not on the population size.



Central Limit Theorem: unknown σ

When the population variance, σ , is unknown, we can use the sample estimate:

$$\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}_x}{\sqrt{n}}$$

$$\hat{\sigma}_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$



Confidence intervals

Instead of an arbitrary threshold it is often more illuminating to present **confidence intervals** or graphical presentations of levels of uncertainty.



Appendix

Probabilities

Probability
distributions
in R

OLS in R
manually

Likelihood
tests

References

Outline

- 4 Probabilities
- 5 **Probability distributions in R**
- 6 OLS in R manually
- 7 Likelihood tests



- You know x and want to know the **density** at that point:
d



Probability distributions in R

- You know x and want to know the **density** at that point:
d
- You know x and want to know the **area** up to that point:
p



- You know x and want to know the **density** at that point:
d
- You know x and want to know the **area** up to that point:
p
- You know x and want to know the area beyond that point:
1-p



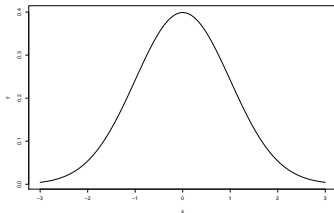
- You know x and want to know the **density** at that point: **d**
- You know x and want to know the **area** up to that point: **p**
- You know x and want to know the area beyond that point: **1-p**
- You know the area and want to know the x value: **q**



- You know x and want to know the **density** at that point:
d
- You know x and want to know the **area** up to that point:
p
- You know x and want to know the area beyond that point:
1-p
- You know the area and want to know the x value: **q**
- You want **random** numbers drawn from that distribution:
r



Probability distributions in R: dnorm

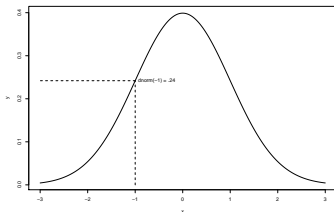


```
x <- seq(-3,3,.01)
```

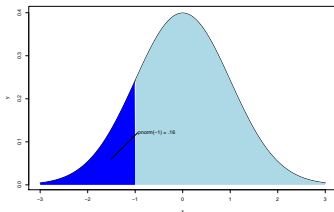
```
y <- dnorm(x)
```



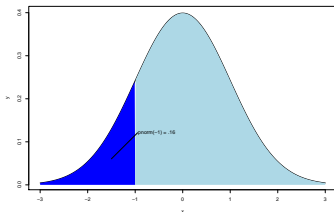
Probability distributions in R: dnorm



```
> dnorm(-1)
[1] 0.2419707
```

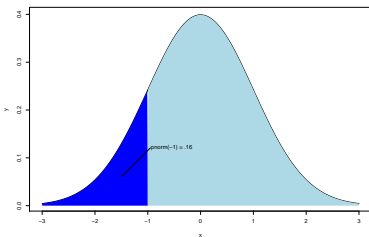


```
> pnorm(-1)
[1] 0.1586553
```

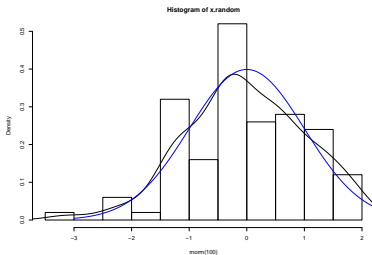


```
> 1 - pnorm(-1)
```

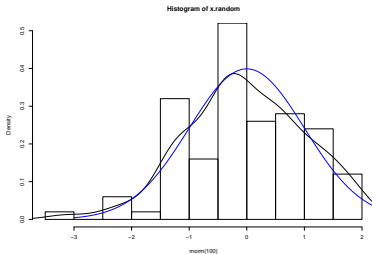
```
[1] 0.8413447
```



```
> qnorm(.1586553)  
[1] -0.9999998
```



```
x.random <- rnorm(100)
```



```
hist(x.random, freq=false)
```



These functions work on many distributions:

<code>rnorm()</code>	random from normal distribution
<code>pchisq()</code>	get area under χ^2 -distribution
<code>1-pf()</code>	get area under F -distribution
<code>rbinom()</code>	draw randomly from binomial distribution
<code>1-dt()</code>	get p -value from t -distribution (one-tailed)



Twenty throws (“Bernoulli trials”) with a coin:

```
> x <- rbinom(20,1,.5)
> factor(x, labels=c("head","tails"))
[1] head  head  head  tails head  tails tails head
[9] tails head  head  head  tails tails tails head
[17] tails tails tails head
```



Outline

- 4 Probabilities
- 5 Probability distributions in R
- 6 OLS in R manually
- 7 Likelihood tests

Appendix

Probabilities

Probability
distributions
in R

OLS in R
manually

Likelihood
tests

References



```
n <- dim(X)[1]
k <- dim(X)[2]
b.hat <- solve(t(X) %*% X) %*% t(X) %*% y
e <- y - X %*% b.hat
s2.hat <- 1/(n-k) * t(e) %*% e
v.hat <- s2.hat %x% solve(t(X) %*% X)
```



t -test

```
bhat <- solve(t(x) %*% x) %*% t(x) %*% y
e <- y - x %*% bhat
vhat <- (1/(n-k) * t(e) %*% e) %x% solve(t(x) %*% x)
se <- sqrt(diag(vhat))
p <- 2 * (1 - pt(abs(bhat / se), n-k))
cbind(bhat, se, p)
```

`abs()` absolute value

`2 *` because it is a two-tailed test



Appendix

Probabilities

Probability
distributions
in R

OLS in R
manually

Likelihood
tests

References

```
ssr <- t(e) %*% e
yhat <- X %*% bhat
sse <- t(yhat - mean(y)) %*% (yhat - mean(y))
F <- (sse/(k-1))/(ssr/(n-k))
1 - pf(F, k-1, n-k)
```



Outline

- 4 Probabilities
- 5 Probability distributions in R
- 6 OLS in R manually
- 7 Likelihood tests

Appendix

Probabilities

Probability
distributions
in R

OLS in R
manually

Likelihood
tests

References



Likelihood

“The likelihood that any parameter (or set of parameters) should have any assigned value (or set of values) is proportional to the probability that if this were so, the totality of observations should be that observed.”

(Fisher, 1922, 310)



Maximum Likelihood

The likelihood is proportional to the probability of observing the data you have (give or take some arbitrarily small deviation), given some parameter estimate:

$$L(\beta|\mathbf{y}, \mathbf{X}) = \alpha P(\mathbf{y}, \mathbf{X}|\beta)$$

The **likelihood function** is thus proportional to the **probability density function** of the given sample, as a function of the parameter values.

The estimator that maximizes this function also maximizes this probability density function and is the **Maximum Likelihood Estimator** (MLE or ML).



Three tests

Three tests are based on this likelihood:

- **Likelihood Ratio (LR) test:**

$$H_0 : 2 \log \frac{L_U}{L_R} = 2(\log L_U - \log L_R) = 0$$

- **Wald (W) test:**

$$H_0 : \frac{(\hat{\beta}^{MLE} - \beta^*)^2}{\text{var}(\hat{\beta}^{MLE})} = 0$$

- **Lagrange Multiplier (LM) test:**

$$H_0 : \frac{\partial \ln L}{\partial \beta} = 0$$

Three tests



Appendix

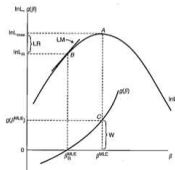
Probabilities

Probability
distributions
in R

OLS in R
manually

**Likelihood
tests**

References





Three tests

Asymptotically, LR, W, and LM are all χ^2 -distributed.

In small samples, $W \geq LR \geq LM$.

We return to these tests and how to do them in R later in the course.

(Kennedy, 2008, 64)



Davidson, Russell and James G. MacKinnon. 1999. *Econometric theory and methods*. Oxford: Oxford University Press.

Fisher, Ronald A. 1922. "On the Mathematical Foundations of Theoretical Statistics." *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 222(594–604):309–368.

URL: <http://rsta.royalsocietypublishing.org/content/222/594-604/309>

Gerber, Alan and Neil Malhotra. 2008. "Do statistical reporting standards affect what is published? Publication bias in two leading political science journals." *Quarterly Journal of Political Science* 3(3):313–326.

Harville, David A. 2008. *Matrix algebra from a statistician's perspective*. New York: Springer-Verlag.

Hayashi, Fumio. 2000. *Econometrics*. Princeton: Princeton University Press.

Kennedy, Peter. 2008. *A guide to econometrics*. 6th ed. Malden, MA: Blackwell.

King, Gary. 1998. *Unifying political methodology. The likelihood theory of statistical inference*. University of Michigan Press.

Wooldridge, Jeffrey M. 2002. *Introductory econometrics: A modern approach*. South-Western.